

TREBALL FI DE GRAU

Grau en Enginyeria de l'Energia

**ESTUDI DEL COMPORTAMENT DE LES CRIPTOMONEDES PER
A LA SEVA POSSIBLE PREDICCIÓ**



Memòria

Autor: Guillem Manresa
Director: Gerard Escudero Bakx
Convocatòria: Juny 2018

Resum

Aquest projecte consisteix en l'estudi del comportament de les criptomonedes, amb la fi de determinar si és viable invertir temps en crec un programa de compra/venta de criptomonedes automàtica.

Si bé ja existeixen altres indicadors que prediuen la tendència d'actius, no s'ha trobat cap indicador fiable que dugui a terme una predicció del preu de l'actiu. Aquest fet es considera necessari per tal de la realització del software de compra/venta automàtica.

Aquest estudi es du a terme en base a la predicció que es fa utilitzant diferents algoritmes de *Machine Learning* ja existents. Per tal d'aplicar aquests algoritmes s'han desenvolupat un seguit d'scripts en el llenguatge de programació Python, que duen a terme l'obtenció de dades relatives a les criptomonedes, indicadors borsaris i divises de diferents països, per posteriorment ser tractades de manera adequada. Després del tractament, les dades estaran llestes per aplicar el procés de *test and training* i dur a terme la predicció del preu de diferents criptomonedes.

Paraules clau: criptomonedes, regressió, predicció, aprenentatge automàtic.

Resumen

Este proyecto consiste en el estudio del comportamiento de las criptomonedas, con la finalidad de determinar si es viable invertir tiempo en la creación de un programa de compra/venta automática de criptomonedas.

Si bien ya existen otros indicadores que predicen la tendencia de activos, no se ha encontrado ningún indicador fiable que lleve a cabo una predicción del precio de dichos activos. Este hecho se considera indispensable para la realización de un software de compra/venta automática.

Este estudio se basa en la predicción hecha usando diferentes algoritmos, mayoritariamente de *Machine Learning*, ya existentes. Para poder aplicar estos algoritmos correctamente se han desarrollado diversos scripts en el lenguaje de programación Python, que permiten la obtención de datos relativos a las criptomonedas, indicadores bursátiles y divisas de diferentes países, para posteriormente ser tratarlos de manera adecuada. Después del tratamiento, los datos estarán listos para aplicar el proceso de *test and training* y llevar a cabo la predicción del precio de diferentes criptomonedas.

Palabras clave: criptomonedas, regresión, predicción, aprendizaje automático.

Abstract

This project consists of the study of the behaviour of the cryptocurrencies. The purpose is to determinate if it is viable to invest time in the creation of a program that buys and sells cryptocurrencies by itself.

Although there are already other indicators that predict the trend of assets, no reliable indicator, that predicts the price of these assets, has been found. This fact is considered essential for the realization of the program described above.

This study is based on the prediction made using different algorithms, most of them are Machine Learning algorithms, that already exist. In order to apply these algorithms, several scripts have been developed in the Python programming language. The process of obtaining the data related to cryptocurrencies, stock exchange indicators and currencies of different countries is allowed by these scripts. These scripts also treat the data in order to prepare them to be used by the previously mentioned algorithms.

Keywords: cryptocurrencies, regression, prediction, machine learning.



Agraïments

En aquest apartat m'agradaria agrair a diferents persones la seva ajuda i aportació a l'hora de realitzar aquest projecte.

Al meu tutor, Gerard Escudero. Gràcies per donar-me l'oportunitat de dur a terme aquest projecte, així com d'ajudar-me a polir la idea inicial i donar-li més profunditat. Gràcies per, a diferència del meu anterior tutor, haver-te mostrat sempre disposat a ajudar-me i per haver estat sempre disponible.

A l'empresa Filtro-Net S.L., i en especial al Sergi Marqueta, per haver-se mostrat tant comprensius sempre que he tingut que anar a la universitat a fer gestions durant l'horari laboral.

Als meus companys Joan Heredia, Marc Campmany i Gerard Espino, donat que la idea tractada en aquest projecte va sorgir d'un projecte en comú entre els quatre, i pel suport i ajuda donats durant aquest quadrimestre.

A tots vosaltres, moltes gràcies



Índex

RESUM	I
RESUMEN	II
ABSTRACT	III
AGRAÏMENTS	V
1. INTRODUCCIÓ	9
1.1. Motivació	9
1.2. Objectius i abast del projecte	10
2. ESTAT DE L'ART	11
2.1. Indicadors	11
2.1.1. MACD	11
2.1.2. Bandes de Bollinger	13
2.1.3. RSI	14
2.2. Algoritmes	15
2.2.1. Regressió lineal	15
2.2.2. K veïns més propers	15
2.2.3. Arbres de decisió	16
2.2.4. Adaptative boosting	17
2.2.5. Màquines de vectors de suport	17
2.3. Recursos	18
2.3.1. CryptoCompare	18
2.3.2. Google finance	18
2.3.3. Forex Python	19
3. DISSENY I TRACTAMENT DE LES DADES	20
3.1. Selecció de criptomonedes	20
3.2. Extracció i tractament de dades	21
4. EXPERIMENTS	25
4.1. Experiment 1	25
4.2. Experiment 2	29
4.3. Experiment 3	33
4.4. Experiment 4	36

4.5. Experiment 5	40
4.6. Resum dels experiments 1-5	43
4.7. Experiment 6	47
5. ANÀLISI DE L'IMPACTE AMBIENTAL	56
CONCLUSIONS	57
Treballs futurs	58
PLANIFICACIÓ	59
PRESSUPOST I/O ANÀLISI ECONÒMICA	60
Pressupost del prototip	60
Pressupost de producció	61
BIBLIOGRAFIA	62

1. Introducció

1.1. Motivació

Les criptomonedes són un mitja digital d'intercanvi. La idea que hi ha darrere d'aquest tipus de divises és el d'un sistema de transferències universal, sense intermediaris i que garanteixi la seguretat dels usuaris d'aquesta. Les principals avantatges d'aquest sistema són la disminució del cost de la transacció (s'elimina l'intermediari), la reducció del temps (a diferència de la resta de sistemes, les transaccions es realitzen en minuts) i s'elimina la necessitat de fer servir agents financers per dur a terme les transaccions.

Tot i les avantatges anteriorment citades, també existeixen certs inconvenients que fan que molta gent sigui reticent a l'ús d'aquest tipus de moneda. Els principals inconvenients són l'anonimat, i el tractar-se d'un sistema totalment descentralitzat i controlat només pels usuaris d'aquest.

El fet de ser un mercat només regulat pels usuaris a donat lloc a que les criptomonedes, lluny de l'idea per la qual van aparèixer, es converteixin en un mitja d'especulació. El mercat de les criptomonedes ha pres molta importància en els últims mesos, especialment en l'últim quadrimestre de 2017, degut als alts beneficis que es poden obtenir en molt poc temps. Va ser al setembre de 2017 quan va sorgir la idea de que en un mercat en un procés d'expansió tant exagerat seria un bon mercat per invertir.

Per tal d'optimitzar aquesta inversió, es vol crear un programa que compri i vengui de manera automàtica. Donat que aquest mercat està obert 24h al dia, tots els dies de l'any, el programa podria estar funcionant sempre. La principal dificultat rau en la predicció dels valors, donat que es tracta d'un mercat molt volàtil. És per això que es pretén fer un estudi del comportament d'aquests actius i veure si és possible fer una bona predicció del seu valor per determinar si és viable invertir temps en el disseny i implementació d'aquest programa.

1.2. Objectius i abast del projecte

L'objectiu d'aquest treball està inclòs en un projecte més gran (creació d'una aplicació de compra/venta automàtica de criptomonedes). Així doncs, l'objectiu del present treball és el de determinar la viabilitat d'aquesta aplicació.

Donada la gran quantitat de criptomonedes existents, i les nombroses tècniques d'aprenentatge automàtic que es fan servir avui en dia, en aquest projecte només s'estudiarà el comportament d'alguna de les monedes existents. De la mateixa manera, per tal de predir el seu valor, s'utilitzaran diferents algorismes d'aprenentatge autònom.

Per tal d'assolir l'objectiu principal, s'hauran d'anar assolint diferents objectius menors tals com:

- Selecció de les criptomonedes a estudiar.
- Selecció de les diferents tècniques d'aprenentatge automàtic a aplicar.
- Extracció de la informació.
- Correcta manipulació de la informació.
- Aplicació de les tècniques d'aprenentatge automàtic al conjunt de dades obtinguts en els primers dos objectius.

2. Estat de l'art

En aquest capítol es comenten alguns indicadors utilitzats per a invertir en el mercat de les criptomonedes, així com els recursos i les eines utilitzades en aquest projecte.

2.1. Indicadors

Per tal de respondre a la pregunta de si és viable invertir temps en crear un programa o aplicació de compra/venta automàtica en el mercat de les criptomonedes, en aquest projecte es pretén estudiar el comportament del mercat. Amb la finalitat de fer-ho, mitjançant tècniques majoritàriament de *Machine Learning*, es creen una sèrie de scripts que prediuen el valor de diferents criptomonedes (indicadors). En aquest apartat es descriuran els principals indicadors existents, així com les diferents eines i fonts utilitzades.

El mercat de les criptomonedes ha guanyat molta importància en els últims mesos, i és per això que han sorgit o s'han adaptat diferents indicadors que s'utilitzaven en altres mercats per tal de predir o aproximar els valors i tendències dels actius.

2.1.1. MACD

El MACD (*Moving Average Convergence Divergence*) [3] és un indicador que busca predir el comportament d'un actiu per tal d'indicar a l'usuari quan és un bon moment per comprar o vendre aquest actiu en funció de la predicció de la tendència d'aquest.

Aquest indicador està compostat per tres components: el MACD, la senyal i l'histograma.

El **MACD** és la diferència entre dues mitges mòbils exponencials de diferent longitud. La primera de les dues es mostra més sensible als moviments del preu a curt termini (12 períodes), mentre que la segona dóna més importància a les variacions del preu a mig termini (26 períodes). La **senyal** és una altra mitja mòbil exponencial de 9 períodes calculada respecte al MACD, que un cop es dibuixa en un gràfic, es va creuant amb el MACD. Per últim, l'**histograma** no és res més que la resta d'aquestes dues corbes.



Figura 2.1.1.1. MACD, senyal i histograma.

Hi ha diferents maneres de fer servir aquest indicador. La més senzilla i menys precisa es tracta de veure quan el MACD es creua amb la senyal. Quan el MACD creua per sobre de la senyal, i mentre es manté per sobre, la tendència és alcista, i per tant es llança una ordre de compra. D'altra banda quan és la senyal la que està per sobre del MACD, s'emet una ordre de venda donat que la tendència de l'actiu és baixista.

Com s'ha comentat anteriorment, només amb aquesta interpretació no s'obtenen resultats òptims, és per això que s'han de combinar amb altre criteris. Un d'aquests criteris és analitzar el comportament del MACD i la senyal al creuar la línia del 0. Seguint el mateix criteri que al principi quan el MACD creua el 0 per sobre de la senyal, s'emet una ordre de compra i si la creua per sota de la senyal, de venda.

Per últim també és interessant estudiar la divergència. Aquesta pot ésser positiva o negativa. Es parla de divergència positiva quan els màxims de l'indicador MACD són cada cop més baixos, mentre que la tendència del preu va a la alça (màxims més alts), en aquest cas es llança una ordre de venda. De manera contrària, la divergència negativa succeeix quan els mínims de l'indicador són cada cop més alts i el preu de l'acció més baixos. En aquest cas l'ordre emesa és de compra.

2.1.2. Bandes de Bollinger

Un altre indicador, també basat en mitges mòbils (simples o exponencials), són les bandes de Bollinger [4]. Majoritàriament és utilitzat per tal de determinar la volatilitat del valor d'una acció. En aquest cas s'acostuma a fer servir una mitja mòbil de 21 períodes respecte al preu de tancament. A l'hora de dibuixar-ho en un gràfic, s'afegeixen dues corbes més que s'obtenen sumant i restant a la mitja mòbil el doble de la seva desviació típica.

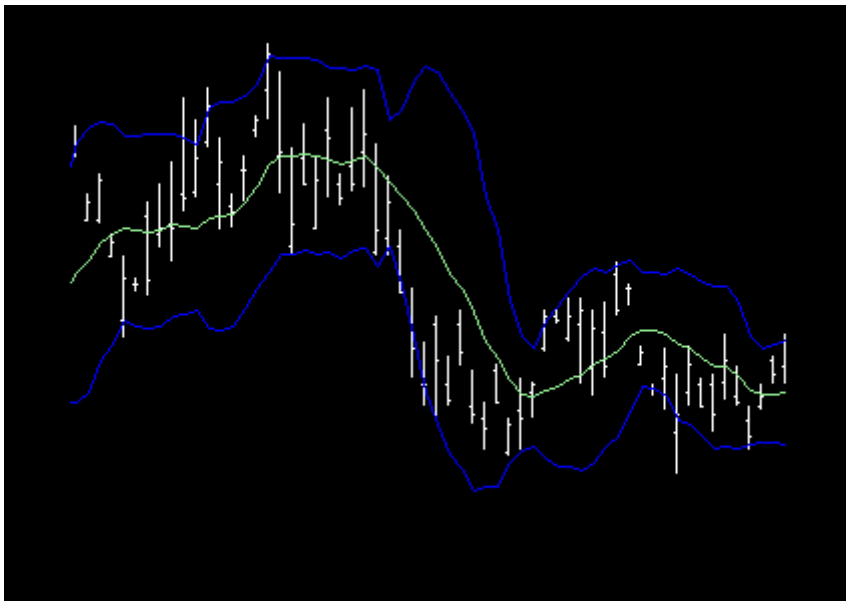


Figura 2.1.2.1. Bandes de Bollinger de 10 períodes amb dos la desviació típica d'amplada.

La interpretació d'aquest indicador és més senzilla que la del MACD. Així doncs, si la distància entre les bandes i la mitja mòbil es redueix, vol dir que la volatilitat de l'actiu és baixa. Contràriament, si aquesta distància s'eixampla, la volatilitat serà més alta.

Una altra manera d'interpretar-lo és parant atenció als preus. Si aquests es troben per sobre de la mitja i són propers a la banda superior, indica que el preu és relativament alt, i que existeix sobre compra. En el cas de que el preu estigui per sota de la mitja i sigui proper a la banda inferior, es pot concloure que el preu és relativament baix i s'està experimentant un període de sobre venda.

2.1.3. RSI

L'índex de força relativa (*Relative Strength Index*) [5] és un altre indicador d'interpretació senzilla, que de la mateixa manera que les bandes de Bollinger, serveix per indicar si un valor està experimentant un període de sobre compra o de sobre venda. L'RSI es basa en el valor dels preus de tancament.

Aquest indicador s'obté calculant un quocient de dues mitges mòbils exponencials. Es calcula per a cada dia la diferència entre el preu de tancament del dia actual i el de l'anterior dia, anomenant U si aquesta diferència és positiva i D si es negativa. Es calcula la mitja mòbil exponencial per a cada un dels dos valors durant el període que es desitja estudiar i el quocient es converteix en un índex que estarà entre 0 i 100.



Figura 2.1.3.1 RSI d'un període de 14 dies.

Com es pot veure en el gràfic, la interpretació d'aquest indicador és molt senzilla. S'estipula que si el RSI es troba per sobre de 70, s'està experimentant sobre compra. El cas contrari, la sobre venda, succeeix quan l'indicador es troba per sota de 30.

2.2. Algoritmes

Per tal de predir el valor de les criptomonedes s'utilitzaran diferents algoritmes ja existents. En el present projecte, s'han fet servir una sèrie d'algoritmes continguts en la llibreria de Python, Scikit-Learn (Sklearn).

La llibreria Sklearn és una de les principals llibreries que existeixen per a treballar amb *Machine Learning*. Aquesta inclou la implementació d'un gran nombre d'algoritmes de classificació, extracció de característiques, regressions, etc. En aquest projecte es faran servir els següents algoritmes continguts en aquesta llibreria:

- Regressió lineal.
- K veïns més propers.
- Arbres de decisió
- *Adaptative boosting*.
- Màquines de vectors de suport.

Es conegut que no existeix un model universal que funcioni millor que la resta per a qualsevol problema o conjunt de dades, és per això que s'han seleccionat aquest algoritmes, donat que fan servir diferent tipus d'informació com a font per al mecanisme d'inducció dels models.

2.2.1. Regressió lineal

Es tracta d'un algoritme més simple que s'utilitzarà en aquest projecte. La regressió lineal pretén estimar valors reals basats en variables contínues. Es tracta d'aproximar la relació de dependència de la variable dependent amb les independents mitjançant una recta. La recta de regressió ve descrita per la següent equació lineal:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (\text{Eq. 2.2.1.1})$$

2.2.2. K veïns més propers

El mètode dels k veïns més propers o kNN és un mètode de classificació supervisada no paramètric que aproxima la probabilitat de que un element pertanyi a una classe en particular mitjançant la informació proporcionada pel conjunt de prototips. En aquest algoritme no es fa cap suposició respecte la distribució de les variables durant el procés d'aprenentatge. La regressió es calcula prenent la mitjana del veí més proper al punt que s'està provant.

Aquest algoritme és bastant susceptible a la presència d'informació irrellevant, tot i que si s'utilitzen valors alts de k, es aconsegueix reduir l'efecte del soroll en la classificació. Donat que amb valors diferents de k es poden obtenir resultats molt diferents, a l'hora d'aplicar aquest algoritme s'optimitzaran tant els valors de k, com la distribució dels pesos per tal d'obtenir el resultat més alt possible.

2.2.3. Arbres de decisió

Aquest algoritme utilitza construccions lògiques molt semblants als sistemes de predicció basats en regles. Aquestes construccions serveixen per categoritzar una sèrie de condicions que succeeixen de manera successiva amb la finalitat trobar la resolució a un problema. Els arbres de decisió estan formats pels següents elements: nodes, vectors, fletxes i etiquetes.

Els **nodes** són els punts en els que s'ha de prendre una decisió d'entre varies. A mesura que augmenten els nodes, van augmentant els possibles finals (nodes terminals). Els **vectors** són la solució final a la que s'arriba en funció de les diferents possibilitats existents. Les **fletxes** són els elements que uneixen els nodes entre ells, representant cada una d'elles, una acció diferent. Finalment, les **etiquetes** donen nom a cada node i fletxa de l'arbre.

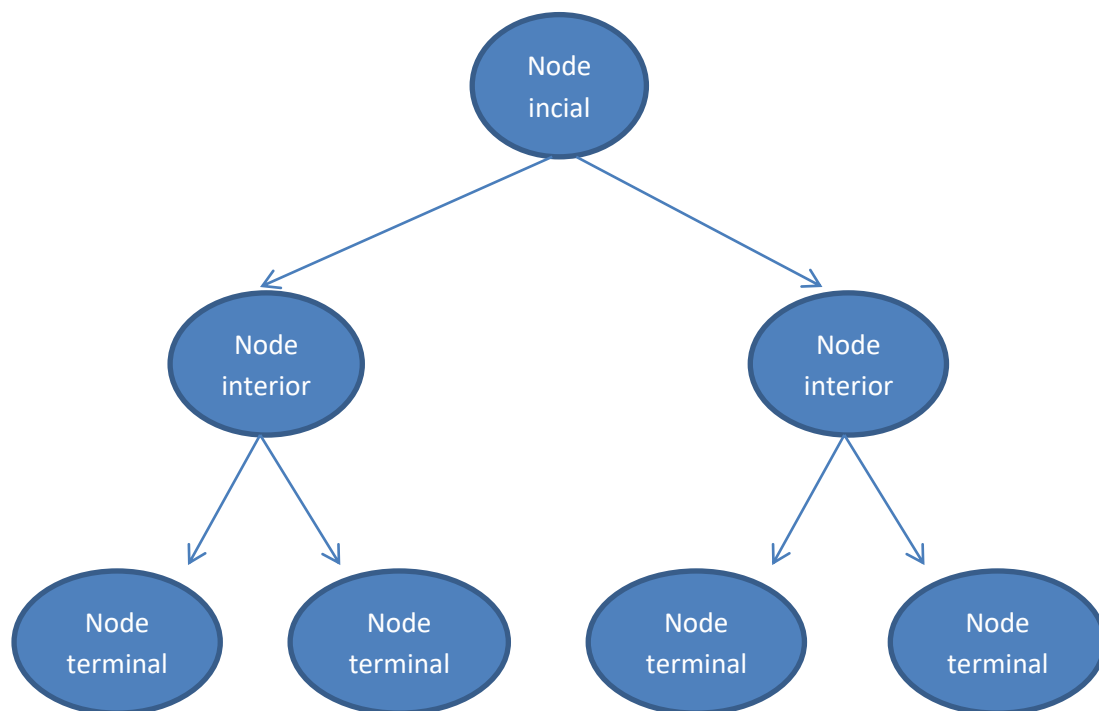


Figura 2.2.3.1. Exemple d'arbre de decisió.

2.2.4. Adaptive boosting

L'algoritme d'*adaptive boosting* no té sentit sense ser utilitzat juntament amb altres algoritmes, ja que el que fa es millorar el rendiment d'un algoritme base. Aquesta millora és duta a terme iterativament al tenir en compte els casos classificats de manera incorrecta en el conjunt d'entrenament. En el present projecte s'utilitzarà l'algoritme basat en arbres de decisió com a algoritme base.

L'algoritme tracta el problema com un cas d'optimització numèrica on l'objectiu és minimitzar la funció de pèrdua. La funció de pèrdua depèn del problema que es tracta. En el cas dels problemes de regressió el més comú es utilitzar l'error quadràtic com a funció a minimitzar.

En el cas dels arbres de decisió es fa servir un model additiu, és a dir, els arbres es van afegint d'un en un, i els arbres ja existents no es canvien. Així doncs es van afegint arbres amb diferents paràmetres de tal manera que la combinació d'aquests minimitzi la pèrdua del model.

Tot i ser un dels models més eficaços i utilitzats en aquest tipus de problemes, aquest mètode té l'inconvenient del temps d'entrenament. Donat que requereix un ajust meticulós dels paràmetres, el temps d'entrenament pot ser molt elevat.

2.2.5. Màquines de vectors de suport

L'idea d'aquest algoritme és trobar un pla que separi els grups de dades en classes. L'algoritme etiqueta les classes de les mostres i entrena una SVM per tal de predir la classe a la que pertanyerà la nova mostra.

Es representen els punts de mostra a l'espai i es separen les classes a dos espais mitjançant un hiperplà definit com el vector entre els dos punts més propers (vector de suport). A l'hora d'aplicar la segregació de classes, sorgeixen diferents problemes, entre ells pot passar que el conjunt de dades no pugui ser completament separat, o que es trobin classificacions en més de dues categories. Per tal d'afrontar aquests inconvenients es fan servir certs paràmetres: C , la funció Kernel.

El paràmetre C controla la compensació entre els errors d'entrenament i els marges rígids, crea un marge que permeti errors en la classificació a l'hora que els penalitza. D'altra banda, la funció Kernel permet canviar l'espai que s'està tractant a un altre de major dimensionalitat.

2.3. Recursos

A l'hora d'obtenir les dades que seran utilitzades en aquest projecte s'han fet servir diferents pàgines webs i llibreries ja existents que proporcionen certa quantitat de dades de manera gratuïta.

2.3.1. CryptoCompare

El recurs que s'ha fet servir per obtenir totes les dades relatives a les criptomonedes ha estat la web CryptoCompare [6]. Aquest lloc web té una interfície de programació d'aplicacions (API) pròpia que permet obtenir les dades d'aproximadament 1800 criptomonedes i de 90 pàgines de compra/venta d'aquestes a temps real.

CryptoCompare retorna una llista de diccionaris que pot ser directament tractada. El cas que ocupa aquest projecte s'ha utilitzat la funcionalitat de l'API que retorna 2000 mostres de la criptomoneda escollida en la pàgina de compra desitjada a partir del febrer de 2015. En cada una d'aquestes mostres apareixen les diferents dades: data, preu d'apertura del període, preu de tancament del període, màxim i mínim durant el període, volum d'inversió del que es parteix a l'inici del període i volum al que s'arriba al finalitzar aquest.

CryptoCompare ofereix les dades per minut, hora o dia. En aquest cas el període seleccionat ha estat el de un dia. Com a inconvenient d'aquest període de temps es troba que moltes de les criptomonedes tenen menys de 2000 dies de vida i no s'aprofita tot el volum de dades que es podria obtenir, però permet donar una visió més general que els altres períodes oferts. En el cas d'obtenir dades cada hora, es tindrien dades d'aproximadament 83 dies, d'altra manera amb el període seleccionat s'està estudiant el comportament d'aquests valors durant més de tres anys en cas d'algunes criptomonedes, fet que permet donar una millor visió del comportament d'aquestes en el temps.

2.3.2. Google finance

Les dades borsàries utilitzades s'obtenen de la llibreria `googlefinance.client`. Es tracta d'una llibreria creada per a Python de l'API de Google Finance [7]. Aquesta llibreria conté les dades dels diferents índex borsaris del món. La llibreria retorna un DataFrame amb les següents dades: els punts del mercat a l'hora de l'apertura, el màxim i el mínim que assoleix el mercat durant el període, els punts del mercat a l'hora del tancament i el volum d'inversió. Així com la pàgina web de CryptoCompare limitava les dades a 2000 mostres, en aquest cas es limiten a 4000 dies de dades.

2.3.3. Forex Python

Per últim, les dades de les diferents monedes de diferents països s'han obtingut de la llibreria Forex Python [8]. Aquesta llibreria permet obtenir, entre d'altres, el registre històric del valor de diferents divises, de cada dia des de l'any 1999. A l'hora, fa servir les dades publicades al Banc Central Europeu, que són actualitzades diàriament. Cal destacar, que a l'hora d'obtenir les dades la limitació d'aquest recurs és més estricta. Per tal de poder obtenir les dades desitjades, s'ha hagut de fer servir la funció de Python `time.sleep()` per tal d'evitar el límit de descàrregues per segon.

3. Disseny i tractament de les dades

En aquest capítol es tracta la selecció de monedes que s'estudiaran al llarg del treball, així com la extracció i tractament que reben les dades.

3.1. Selecció de criptomonedes

Actualment existeixen centenars de criptomonedes, és per això, que a l'hora de dur a terme aquest projecte, s'han hagut de seleccionar algunes de les criptomonedes del mercat, sent impossible estudiar-les totes. A continuació es detalla quines monedes s'han seleccionat, així com el motiu de la selecció.

Donat el gran volum de criptomonedes existents, el primer criteri emprat en la selecció ha sigut el del volum d'inversió. S'ha volgut estudiar el comportament de les monedes més importants, o si més no, les quals mouen més diners. Totes les monedes estudiades en aquest projecte es troben entre les 10 primeres en aquest aspecte.





















#	Name	Market Cap	Price	Volume (24h)	Circulating Supply	Change (24h)	Price Graph (7d)
1	 Bitcoin	\$131.958.180.537	\$7.729,30	\$4.981.860.000	17.072.462 BTC	1,19%	
2	 Ethereum	\$61.917.236.545	\$620,13	\$1.923.790.000	99.845.414 ETH	5,01%	
3	 Ripple	\$26.184.692.953	\$0,667270	\$312.387.000	39.241.525.848 XRP *	3,72%	
4	 Bitcoin Cash	\$20.285.667.963	\$1.181,91	\$844.259.000	17.163.463 BCH	10,13%	
5	 EOS	\$13.301.815.757	\$14,84	\$2.375.760.000	896.149.492 EOS *	1,15%	
6	 Litecoin	\$7.188.682.236	\$126,52	\$328.079.000	56.816.748 LTC	3,29%	
7	 Cardano	\$6.000.768.622	\$0,231448	\$121.989.000	25.927.070.538 ADA *	3,11%	
8	 Stellar	\$5.690.560.550	\$0,306282	\$64.504.900	18.579.480.837 XLM *	2,65%	
9	 IOTA	\$5.147.995.832	\$1,85	\$141.053.000	2.779.530.283 MIOTA *	-1,75%	
10	 TRON	\$4.113.208.439	\$0,062560	\$267.256.000	65.748.111.645 TRX *	1,90%	

Figura 3.1.1. Top 10 de criptomonedes 03/06/2018

Les monedes seleccionades per a ser estudiades han estat les següents: Bitcoin, Litecoin, Ethereum, Ripple i Iota.

El **Bitcoin** ha sigut seleccionat per a ser estudiat degut a varies raons. Es tracta de la primera criptomoneda descentralitzada que existeix i actualment, la resta de les criptomonedes es veuen influenciades pel comportament del Bitcoin. A més a més, es tracta de la moneda amb més volum d'inversió en el mercat actualment.

Tot i estar en sisena posició en quant a volum d'inversió es refereix, el **Litecoin** s'estudia en aquest projecte degut a la seva antiguitat. Tot i no ser la primera criptomoneda que es coneix, va aparèixer el 2011 (dos anys més tard que el Bitcoin) i és la primera en utilitzar un algoritme diferent al del Bitcoin.

L'**Ethereum** és actualment la segona moneda amb més inversió tot i entrar al mercat a l'any 2015. També és molt interessant el fet de que es tracta de la primera moneda que incorpora un sistema de contractes intel·ligents basat en la cadena de blocs (*blockchain*¹).

El **Ripple** és la moneda que ocupa el tercer lloc en la llista que es pot veure en la figura 3.1.1, completant així el top 3 de monedes amb inversió. Cal destacar, que a diferència de les tres monedes anteriorment comentades, les transaccions de Ripple no es validen amb el procés conegut com a mineria².

Tot i estar més endarrerida pel que fa referència a volum d'inversió, la **Iota** és una moneda molt interessant de ser estudiada. De la mateixa manera que l'Ethereum, es tracta d'una plataforma que també incorpora els contractes intel·ligents. A diferència de la resta de criptomonedes, que es basen en la *blockchain*, Iota utilitza una estructura de dades anomenada *tangle*³. Aquesta estructura la dota d'una escalabilitat molt més alta que la resta de monedes.

3.2. Extracció i tractament de dades

La selecció, extracció i tractament de les dades juga un paper molt important en aquest projecte, és per això que s'han estudiat diferents recursos i pàgines web d'on es poden obtenir diferents tipus de dades.

Pel que fa a les criptomonedes, el recurs seleccionat ha estat la pàgina web CryptoCompare. El principal motiu de la selecció d'aquest recurs ha sigut la facilitat d'obtenir un gran volum de dades de manera gratuïta i en un format senzill per treballar-hi. La gran avantatge d'aquesta pàgina és que disposa d'una API gratuïta d'on es poden extreure dades de més de 1800 criptomonedes per a cada pàgina d'intercanvi diferent.

¹ Es tracta d'una estructura de dades en la que la informació s'emmagatzema i es valida en blocs. Els blocs es van agrupant en ordre cronològic.

² Aquest terme fa referència al procés que han de patir els blocs (cada bloc és una transacció) per tal de ser validats per la xarxa.

³ Sistema de validació de les transaccions en el que cada transacció valida les dues anteriors.

Un altre gran avantatge és el format en el que s'extreuen es dades. Es tracta d'una llista de diccionaris, on cada diccionari pertany a una mostra, fent així molt senzilla la seva manipulació amb Python.

En el relatiu a les dades de les diferents borses, s'ha utilitzat Google Finance. El principal motiu és l'accessibilitat a les dades, i que aquestes són públiques. És cert no disposa de la mateixa informació per a totes les borses, sent les europees i americanes més completes que les asiàtiques. Les dades són retornades en un *DataFrame*⁴, format que a priori semblava més complicat de manipular, però un cop s'aprèn a fer servir, és tracta d'un format molt robust i que permet treballar amb fitxers diferents i guardar-los en el format adequat per a aplicar-hi els algorismes de regressió.

Finalment, l'últim recurs d'obtenció de dades ha estat la API de Forxe-Python, d'on s'han obtingut les dades relatives a les monedes de diferents països. Cal destacar que tot i tractar-se de dades públiques, el recurs limita les dades que es poden extreure per segon. Inicialment suposava un gran problema, degut que la quantitat de dades que es podien obtenir era molt inferior a les necessàries, però finalment es va poder resoldre el problema afegint un *delay* a l'script. Les dades són obtingudes en forma de diccionari, retornant un diccionari amb el valor de 32 monedes diferents per a una data seleccionada.

Un cop s'està en disposició de totes les dades que han d'involucrar-se en la predicció és necessari ajuntar-les en únic arxiu de manera coherent i ordenada. Aquí és on han aparegut gran part de les dificultats.

Prèviament a ajuntar tota la informació en un únic arxiu s'han hagut de tractar per separat per deixar-los en el mateix format. És per això que s'ha decidit treballar amb la estructura de dades *pandas*⁵. La principal avantatge d'aquesta estructura es que permet ajuntar grans quantitats d'informació i agrupar-la i ordenar-la per índexs. En aquest cas, es fa servir com a índex la data, donat que tots els atributs estan relacionats pel temps.

L'arxiu amb la informació de les criptomonedes ha sigut bastant senzill de tractar, l'única particularitat ha estat modificar el format en el que es trobava la data, donat que es trobava en el format *timestamp*⁶.

⁴ Estructura bidimensional de dades.

⁵ Biblioteca de software creada per a la manipulació i anàlisi de dades.

⁶ Seqüència de caràcters que representen una data específica expressada en segons.

time	close	high	low	open	volumefrom	volumeto
1424044800	236.78	243.65	228.62	234.14	45750.85	10798751.12
1424131200	242.87	246.28	231.5	236.78	42669.08	10228035.45
1424217600	234.8	244.99	231.01	242.87	38588.38	9164141.83
1424304000	241.9	243.42	234.62	234.8	22792.55	5446200.04
1424390400	245.53	248.98	238.95	241.9	33337.97	8136342.67
1424476800	245.1	247.73	243.28	245.53	16514.04	4052466.26
1424563200	236.59	247.78	232.1	245.1	33394.5	7971863.56

Figura 3.2.1. Fragment de l'arxiu .csv⁷ que conté la informació relativa a les criptomonedes.

En la figura 3.2.1 es mostra un fragment de les dades relatives a la informació extreta de les criptomonedes. Les diferents columnes contenen la data, el preu al que tanca el dia, el màxim assolit durant el dia, el mínim assolit durant el dia, el preu comença el dia, el volum de transaccions en la criptomoneda i el volum de transaccions referenciat al dolar americà. Totes les columnes a excepció de la data i el 'volumefrom' estan referenciades al dolar americà.

L'arxiu amb les dades de les borses no s'ha hagut de tractar, donat que s'obté directament com a *DataFrame*.

time	Open	High	Low	Close	Volume	Open.1	High.1	Low.1	Close.1	Volume.1	Open.2
09/05/2014	16551.23	16588.77	16498.71	16583.34	75016850.0	10573.0	10592.6	10438.7	10487.2	324681592.0	14101.98
12/05/2014	16584.82	16704.84	16584.82	16695.47	71553669.0	10512.4	10578.2	10478.2	10567.0	243557039.0	14173.45
13/05/2014	16695.92	16735.51	16695.92	16715.44	71336153.0	10587.3	10600.9	10533.4	10587.2	272278898.0	14367.94
14/05/2014	16716.08	16717.56	16595.0	16613.97	71969085.0	10602.8	10639.1	10521.6	10613.9	248096723.0	14385.6
15/05/2014	16613.52	16622.9	16397.46	16446.81	106617004.0	10616.5	10665.1	10323.4	10365.0	409937852.0	14280.51
16/05/2014	16447.32	16498.99	16414.32	16491.31	119326761.0	10354.1	10494.9	10282.4	10478.7	312919019.0	14087.76
19/05/2014	16490.35	16526.26	16442.12	16511.86	82184256.0	10469.7	10526.0	10307.3	10425.5	243528908.0	14130.05

Figura 3.2.2. Fragment de l'arxiu .csv que conté la informació relativa a les borses.

En la figura 3.2.2 es mostra un fragment de les dades relatives a la informació extreta de les diferents borses. No es poden apreciar totes les columnes de l'arxiu, però elles contenen: la data, els punts de la borsa a l'inici del dia, el màxim assolit durant el dia, el mínim assolit durant el dia, els punts de la borsa al final del dia i el volum de transaccions durant el dia per a cada borsa. Totes les dades a excepció de la data estan referenciades al dolar americà.

Finalment, per a tenir la informació de les 32 monedes implicades, s'ha hagut de dur a terme un procés iteratiu per a obtenir totes les dades desitjades, ja que el recurs proporciona les dades per a una data concreta. Cal destacar que l'extracció d'aquestes dades és molt lenta (15 minuts aproximadament), ja que després de l'obtenció d'una mostra s'atura el procés 0.5 segons per tal d'evitar la limitació de dades/segon.

⁷ Es tracta d'un tipus d'arxiu (*Coma Separated Values*) en el que les dades estan separades per comes o algun caràcter específic.

time	AUD	BGN	BRL	CAD	CHF	CNY	CZK	DKK	EUR
10/02/2015	12.855	17.313	28.008	12.471	0.92564	62.417	24.544	65.894	0.88519
11/02/2015	1.291	17.287	28.611	12.611	0.92496	62.427	24.465	65.795	0.88386
12/02/2015	13.031	17.265	28.747	1.256	0.93212	62.438	24.454	65.718	0.88277
13/02/2015	12.907	17.185	28.464	12.508	0.92927	62.394	24.286	65.407	0.87866
14/02/2015	12.907	17.185	28.464	12.508	0.92927	62.394	24.286	65.407	0.87866
15/02/2015	12.907	17.185	28.464	12.508	0.92927	62.394	24.286	65.407	0.87866
16/02/2015	12.859	17.144	28.345	12.453	0.93145	62.493	24.227	65.252	0.87658
17/02/2015	12.784	17.134	28.379	12.374	0.93132	62.552	24.195	65.212	0.87604

Figura 3.2.3. Fragment de l'arxiu .csv que conté la informació relativa a les monedes.

En la figura 3.2.3 es mostra un fragment de les dades relatives a la informació extreta de les diferents monedes. No es poden apreciar totes les monedes contingudes en l'arxiu, però les columnes contenen la data i el valor de les diferents monedes referenciades al dolar americà per ordre alfabètic.

Un cop en disposició dels tres arxius en el mateix format, simplement han hagut d'ajuntar-se en funció de les necessitats de cada experiment, creant així un arxiu de dades concret per a cada experiment.

4. Experiments

En aquest capítol apareixen els diferents experiments realitzats durant el treball que tenen com a objectiu el de predir el comportament de les criptomonedes seleccionades.

4.1. Experiment 1

En aquest experiment s'utilitzaran com a atributs única i exclusivament dades relatives a les criptomonedes. El conjunt de dades està format pels següents valors: el valor de la criptomoneda al acabar el període de temps seleccionat, el valor de la criptomoneda al iniciar el període, el valor màxim i el mínim de la criptomoneda durant el mateix període, el volum de transaccions en la criptomoneda i el volum de transaccions referenciat al dolar americà. En el cas que segueix s'obtenen dades de manera diària, és a dir, els períodes considerats són de 24 hores.

D'aquestes dades es seleccionen les següents com a atributs per tal de calcular el valor de la criptomoneda: valor màxim de la moneda, valor mínim de la moneda, valor de la moneda a l'inici del període, el volum de transaccions en la criptomoneda i el volum de transaccions referenciat al dolar.

Depenent de la criptomoneda analitzada el conjunt de dades serà més o menys extens, ja que no totes les monedes van sorgir al mateix moment.

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Regressió lineal (LR)
- K veïns més propers (kNN)
- Arbres de decisió (DT)
- *Adaptive boosting* (AB)
- Màquines de vectors de suport (SVM)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les dades de manera aleatòria els valors d' R^2 obtinguts variaran. A les taules apareixerà també el temps d'execució de cada algoritme i els paràmetres que aquest ha utilitzat (en el cas que sigui necessari) per tal d'optimitzar el resultat.

El conjunt d'entrenament estarà format per aproximadament el 60% de les dades, i es validarà amb les dades restants.

En les taules 3.1.1 – 3.1.5, es poden observar els resultats obtinguts separats per criptomoneda:

Bitcoin	R^2		Temps (s)	Paràmetres
LR	0.998	± 0.0005	0.0186	N/A
kNN	0.863	± 0.0175	0.2792	K={11}, pesos={uniforme}
DT	0.995	± 0.0008	0.0208	N/A
AB	0.992	± 0.0008	2.0742	25
SVM	-0.180	± 0.0220	7.3572	kernel=rbf

Taula 3.1.1. 1185 dades amb un grup d'entrenament de 700 dades.

Litecoin	R^2		Temps (s)	Paràmetres
LR	0.998	± 0.0004	0.0188	N/A
kNN	0.801	± 0.0203	0.2922	K={11}, pesos={distància}
DT	0.993	± 0.0033	0.0208	N/A
AB	0.992	± 0.0019	1.9702	25
SVM	0.014	± 0.0014	8.2002	kernel=rbf

Taula 3.1.2. 1186 dades amb un grup d'entrenament de 700 dades.

Ethereum	R^2		Temps (s)	Paràmetres
LR	0.998	± 0.0000	0.0156	N/A
kNN	0.777	± 0.0303	0.1720	K={11}, pesos={uniforme}
DT	0.993	± 0.0015	0.0166	N/A
AB	0.991	± 0.0020	1.4118	50
SVM	0.001	± 0.0077	2.8864	kernel=rbf

Taula 3.1.3. 791 dades amb un grup d'entrenament de 450 dades.

Ripple	R^2		Temps (s)	Paràmetres
LR	0.993	± 0.0005	0.0404	N/A
kNN	0.925	± 0.0316	0.1446	K={3}, pesos={distància}
DT	0.979	± 0.0082	0.0414	N/A
AB	0.978	± 0.0067	1.5458	75, 100
SVM	-0.017	± 0.0370	0.7240	kernel=rbf

Taula 3.1.4. 355 dades amb un grup d'entrenament de 200 dades.

Iota	R^2		Temps (s)	Paràmetres
LR	0.993	± 0.0018	0.0146	N/A
kNN	0.952	± 0.0122	0.1256	K={3, 5}, pesos={uniforme, distància.}
DT	0.981	± 0.0095	0.0152	N/A
AB	0.980	± 0.0102	1.5034	100
SVM	-0.006	± 0.0084	0.7326	kernel=rbf

Taula 3.1.5. 330 dades amb un grup d'entrenament de 200 dades.

Com es pot observar en les diferents taules, per totes les criptomonedes, els valors més alts de l' R^2 s'obtenen utilitzant els algoritmes LR, DT i AB. Per saber si la diferència entre els aquests mètodes és significativa s'hauria de recórrer a algun test estadístic. Es pot observar que els valors obtinguts són més baixos per a les monedes Ripple i Iota. Aquest fet es pot atribuir al menor volum de dades.

Cal destacar també els valors de l'algoritme SVM. En les cinc monedes estudiades els valors de l' R^2 obtinguts utilitzant aquest algoritme són molt propers a 0, agafant inclús valors negatius.

En la columna de paràmetres es poden veure els paràmetres utilitzats pels algoritmes més repetits durant els experiments. En cada experiment s'han optimitzat els diferents paràmetres de cada algoritme per tal d'obtenir el valor d' R^2 més alt possible. El fet de que els paràmetres no siguin constants en cada experiment indica que, si fossin fixats, els resultats serien més baixos.

Així doncs, com a conclusió d'aquest experiment, es pot determinar que l'algoritme que millor funciona és el de regressió lineal. Aquest fet pot indicar que el model que s'està estudiant es massa simple.

4.2. Experiment 2

En aquest experiment s'utilitzaran com a atributs dades relatives a les criptomonedes i a borses de diferents països. El conjunt de dades està format pels mateixos valors relatius a les criptomonedes que en el primer experiment, però afegint informació de 9 indicadors borsaris diferents tals com:

- Dow Jones Industrial Average
- Ibex 35
- Nikkei 225 (índex borsari de Tokio)
- Nasdaq
- Topix (índex borsari de Tokio)
- FTSE (índex borsari de Londres)
- SSE Composite Index (índex borsari de Xangai)
- DAX Performance Index (índex borsari d'Alemanya)
- BSE SENSEX (índex borsari de Bombai)

D'aquests diferents índexs s'obtenen les següents dades: valor màxim i mínim durant el període de temps tractat, valor d'apertura i tancament del període de temps tractat i el volum total d'inversió del període de temps tractat. En el cas que segueix s'obtenen dades de manera diària, és a dir, els períodes considerats són de 24 hores. Cal destacar que els diferents índex borsaris, a diferència de les criptomonedes, tanquen diversos dies a l'any. Per tal de no deixar valors buits, fet que impediria que l'algoritme funcione correctament, s'ha assignat el valor de -1 a les caselles buides dels dies que les borses romanen tancades.

D'aquestes dades es seleccionen com a atributs les mateixes que en el primer experiment afegint-hi totes les dades dels 9 índexs borsaris.

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Regressió lineal (LR)
- K veïns més propers (kNN)
- Arbres de decisió (DT)
- *Adaptive boosting* (AB)
- Màquines de vectors de suport (SVM)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les

dades de manera aleatòria els valors d' R^2 obtinguts variaran. El conjunt d'entrenament estarà format per aproximadament el 60% de les dades, i es validarà amb les dades restants.

En les taules 3.2.1 – 3.2.5, es poden observar els resultats obtinguts separats per criptomoneda:

Bitcoin	R^2		Temps (s)	Paràmetres
LR	0.997	± 0.0029	0.0988	N/A
kNN	0.874	± 0.0145	1.1220	$K=\{11, 7\}$, pesos={uniforme, distància}
DT	0.995	± 0.0008	0.0382	N/A
AB	0.992	± 0.0014	5.7850	75, 100
SVM	-0.163	± 0.0312	13.3380	kernel=rbf

Taula 3.2.1. 1185 dades amb un grup d'entrenament de 700 dades.

Litecoin	R^2		Temps (s)	Paràmetres
LR	0.995	± 0.0037	0.0216	N/A
kNN	0.774	± 0.0320	1.1016	$K=\{11\}$, pesos={uniforme, distància.}
DT	0.993	± 0.0007	0.0384	N/A
AB	0.994	± 0.0008	4.3120	50, 25
SVM	-0.002	± 0.0040	13.3830	kernel=rbf

Taula 3.2.2. 1186 dades amb un grup d'entrenament de 700 dades.

Ethereum	R^2		Temps (s)	Paràmetres
LR	0.974	± 0.0297	0.0200	N/A
kNN	0.791	± 0.0363	0.5602	K={11}, pesos={uniforme, distància.}
DT	0.991	± 0.0020	0.0282	N/A
AB	0.990	± 0.0008	3.4432	100
SVM	-0.003	± 0.0041	5.6414	kernel=rbf

Taula 3.2.3. 791 dades amb un grup d'entrenament de 450 dades.

Ripple	R^2		Temps (s)	Paràmetres
LR	0.962	± 0.0651	0.0200	N/A
kNN	0.807	± 0.0273	0.7130	K={3, 5}, pesos={distància}
DT	0.974	± 0.0103	0.0206	N/A
AB	0.975	± 0.0069	2.7324	50, 25
SVM	-0.004	± 0.0033	1.2844	kernel=rbf

Taula 3.2.4. 355 dades amb un grup d'entrenament de 200 dades.

Iota	R^2		Temps (s)	Paràmetres
LR	0.472	± 1.0739	0.0194	N/A
kNN	0.717	± 0.0339	0.2850	K={3, 5}, pesos={distància}
DT	0.983	± 0.0082	0.0218	N/A
AB	0.983	± 0.0068	2.8634	100, 75
SVM	-0.004	± 0.0024	1.3312	kernel=rbf

Taula 3.2.5. 330 dades amb un grup d'entrenament de 200 dades.

Com es pot observar a les taules dels resultats, a diferència del primer experiment, l'algoritme de regressió lineal no dona els valors més als d' R^2 . A més a més, en les criptomonedes de les quals es disposen menys dades aquest algoritme comença a no funcionar correctament, això pot ser degut a l'augment de complexitat del model.

En la columna de paràmetres es poden veure els paràmetres utilitzats pels algoritmes més repetits durant els experiments. En cada experiment s'han optimitzat els diferents paràmetres de cada algoritme per tal d'obtenir el valor d' R^2 més alt possible. El fet de que els paràmetres no siguin constants en cada experiment indica que, si fossin fixats, els resultats serien més baixos.

D'altra banda es pot observar que, tot i afegir complexitat al model, l'algoritme SVM segueix funcionant igual de malament que en el primer experiment.

Comparant els resultats obtinguts amb els del primer experiment es pot observar que a banda dels canvis en les monedes amb menys dades mencionats anteriorment, la resta segueix donant valors molt semblants.

4.3. Experiment 3

En aquest experiment s'utilitzaran com a atributs dades relatives a les criptomonedes i a monedes de diferents països. El conjunt de dades està format pels mateixos valors relatius a les criptomonedes que en el primer experiment, però afegint-hi el valor de 32 monedes de diferents països del món.

D'aquestes dades es seleccionen com a atributs les mateixes que en el primer experiment afegint-hi les dades de les 32 monedes.

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Regressió lineal (LR)
- K veïns més propers (kNN)
- Arbres de decisió (DT)
- *Adaptive boosting* (AB)
- Màquines de vectors de suport (SVM)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les dades de manera aleatòria els valors d' R^2 obtinguts variaran. El conjunt d'entrenament estarà format per aproximadament el 60% de les dades, i es validarà amb les dades restants.

En les taules 3.3.1 – 3.3.5, es poden observar els resultats obtinguts separats per criptomoneda:

Bitcoin	R^2		Temps (s)	Paràmetres
LR	0.998	± 0.0004	0.0220	N/A
kNN	0.858	± 0.0206	0.5530	K={11}, pesos={uniforme, distància}
DT	0.994	± 0.0008	0.0344	N/A
AB	0.992	± 0.0008	4.4390	50, 75
SVM	-0.164	± 0.0135	11.8136	kernel=rbf

Taula 3.3.1. 1172 dades amb un grup d'entrenament de 700 dades.

Litecoin	R^2		Temps (s)	Paràmetres
LR	0.998	± 0.0000	0.0210	N/A
kNN	0.779	± 0.0271	0.5468	$K=\{7, 11\}$, pesos={uniforme, distància}
DT	0.993	± 0.0015	0.0376	N/A
AB	0.992	± 0.0019	3.6914	25, 50
SVM	0.004	± 0.0033	11.7352	kernel=rbf

Taula 3.3.2. 1172 dades amb un grup d'entrenament de 700 dades.

Ethereum	R^2		Temps (s)	Paràmetres
LR	0.997	± 0.0004	0.0190	N/A
kNN	0.769	± 0.0146	0.3296	$K=\{11\}$, pesos={uniforme}
DT	0.991	± 0.0028	0.0260	N/A
AB	0.991	± 0.0007	3.1590	75, 25
SVM	0.001	± 0.0083	4.5746	kernel=rbf

Taula 3.3.3. 778 dades amb un grup d'entrenament de 450 dades.

Ripple	R^2		Temps (s)	Paràmetres
LR	0.992	± 0.0013	0.0170	N/A
kNN	0.928	± 0.0327	0.2212	$K=\{3, 1\}$, pesos={uniforme, distància.}
DT	0.957	± 0.0209	0.0198	N/A
AB	0.971	± 0.0177	2.4244	10, 50
SVM	-0.005	± 0.0070	1.0336	kernel=rbf

Taula 3.3.4. 342 dades amb un grup d'entrenament de 200 dades.

lota	R^2		Temps (s)	Paràmetres
LR	0.991	± 0.0033	0.0198	N/A
kNN	0.948	± 0.0164	0.2196	$K=\{3, 5\}$, pesos={uniforme, distància}
DT	0.980	± 0.0078	0.0630	N/A
AB	0.980	± 0.0087	2.5198	100, 75
SVM	-0.031	± 0.0341	1.0882	kernel=rbf

Taula 3.3.5. 317 dades amb un grup d'entrenament de 200 dades.

En les taules de resultats es pot observar que, de la mateixa manera que en el primer experiment, els algoritmes que millor funcionen per a les monedes Bitcoin, Litecoin i Ethereum són els mateixos: LR, DT i AB.

D'altra banda, quan s'estudien les dues monedes restants (Ripple i lota), es pot veure com els valors d' R^2 disminueixen per als algoritmes DT i AB, mentre que la regressió lineal segueix donant els mateixos resultats que amb les monedes que disposen de més dades.

Amb l'algoritme SVM es segueixen obtenint resultats semblants i igual de dolents que en els dos primers experiments, fet que sembla indicar que en el model que s'està treballant no serà possible trobar un model que funcioni correctament emprant plans.

En la columna de paràmetres es poden veure els paràmetres utilitzats pels algoritmes més repetits durant els experiments. En cada experiment s'han optimitzat els diferents paràmetres de cada algoritme per tal d'obtenir el valor d' R^2 més alt possible. El fet de que els paràmetres no siguin constants en cada experiment indica que, si fossin fixats, els resultats serien més baixos.

Comparant els resultats amb de l'experiment anterior sembla ser que el valor de les diferents monedes té més influència en el preu de les criptomonedes que les dels índex bursàtils, donat que els resultats són una mica més alts en aquest experiment. Per tal d'acabar de corroborar-ho, en el següent experiment s'intentaran millorar les dades de la borsa substituint els valors buits (-1) per l'últim valor conegut.

4.4. Experiment 4

En aquest experiment s'utilitzaran com a atributs dades relatives a les criptomonedes i a borses de diferents països. El conjunt de dades està format pels mateixos valors relatius a les criptomonedes que en el primer experiment, però afegint informació de 9 indicadors borsaris diferents tals com:

- Dow Jones Industrial Average
- Ibex 35
- Nikkei 225 (índex borsari de Tokio)
- Nasdaq
- Topix (índex borsari de Tokio)
- FTSE (índex borsari de Londres)
- SSE Composite Index (índex borsari de Xangai)
- DAX Performance Index (índex borsari d'Alemanya)
- BSE SENSEX (índex borsari de Bombai)

D'aquests diferents índexs s'obtenen les mateixes dades que de les criptomonedes, és a dir: valor màxim i mínim durant el període de temps tractat, valor d'apertura i tancament del període de temps tractat i el volum total d'inversió del període de temps tractat. En el cas que segueix s'obtenen dades de manera diària, és a dir, els períodes considerats són de 24 hores. Cal destacar que els diferents índex borsaris, a diferència de les criptomonedes, tanquen diversos dies a l'any. A diferència del segon experiment, on els valors buits eren substituïts per -1, en aquest experiment es substituiran pels valors del dia anterior, considerant que els dies que els índexs estan tancats es mantindran constants els valors als que es va arribar l'últim dia operatiu.

D'aquestes dades es seleccionen com a atributs les mateixes que en el segon experiment.

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Regressió lineal (LR)
- K veïns més propers (kNN)
- Arbres de decisió (DT)
- *Adaptive boosting* (AB)
- Màquines de vectors de suport (SVM)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les

dades de manera aleatòria els valors d' R^2 obtinguts variaran. El conjunt d'entrenament estarà format per, aproximadament, el 60% de les dades, i es validarà amb les dades restants.

En les taules 3.4.1 – 3.4.5, es poden observar els resultats obtinguts separats per criptomoneda:

Bitcoin	R^2		Temps (s)	Paràmetres
LR	0.997	± 0.0019	0.0196	N/A
kNN	0.892	± 0.0208	1.0106	K={7, 5}, pesos={distància}
DT	0.994	± 0.0008	0.0384	N/A
AB	0.992	± 0.0014	5.2320	50, 25
SVM	-0.186	± 0.0462	12.4364	kernel=rbf

Taula 3.4.1. 1185 dades amb un grup d'entrenament de 700 dades.

Litecoin	R^2		Temps (s)	Paràmetres
LR	0.997	± 0.0004	0.0182	N/A
kNN	0.865	± 0.0583	1.0838	K={7, 5}, pesos={distància}
DT	0.990	± 0.0023	0.0310	N/A
AB	0.994	± 0.0021	3.5976	25, 50
SVM	-0.002	± 0.0043	12.0714	kernel=rbf

Taula 3.4.2. 1186 dades amb un grup d'entrenament de 700 dades.

Ethereum	R^2		Temps (s)	Paràmetres
LR	0.995	± 0.0053	0.0206	N/A
kNN	0.862	± 0.0082	0.5652	K={7}, pesos={distància}
DT	0.992	± 0.0025	0.0310	N/A
AB	0.990	± 0.0014	4.1636	10, 100
SVM	-0.003	± 0.0068	5.7266	kernel=rbf

Taula 3.4.3. 791 dades amb un grup d'entrenament de 450 dades.

Ripple	R^2		Temps (s)	Paràmetres
LR	0.922	± 0.1009	0.0158	N/A
kNN	0.839	± 0.0519	0.2638	K={3, 5}, pesos={distància}
DT	0.964	± 0.0229	0.0158	N/A
AB	0.982	± 0.0028	2.4476	10, 100
SVM	-0.006	± 0.0055	1.1128	kernel=rbf

Taula 3.4.4. 355 dades amb un grup d'entrenament de 200 dades.

Iota	R^2		Temps (s)	Paràmetres
LR	0.771	± 0.3092	0.0188	N/A
kNN	0.809	± 0.0800	0.2654	K={3, 5}, pesos={distància}
DT	0.977	± 0.0116	0.0214	N/A
AB	0.982	± 0.0028	2.4476	50, 100
SVM	-0.006	± 0.0055	1.1128	kernel=rbf

Taula 3.4.5. 330 dades amb un grup d'entrenament de 200 dades.

	Bitcoin		Litecoin		Ethereum		Ripple		lota	
	Exp.2	Exp.4	Exp.2	Exp.4	Exp.2	Exp.4	Exp.2	Exp.4	Exp.2	Exp.4
LR	0.997	0.997	0.995	0.997	0.974	0.995	0.962	0.922	0.472	0.771
kNN	0.874	0.892	0.774	0.865	0.791	0.862	0.807	0.839	0.717	0.809
DT	0.995	0.994	0.993	0.990	0.991	0.992	0.974	0.964	0.983	0.977
AB	0.992	0.992	0.994	0.994	0.990	0.990	0.975	0.982	0.983	0.982
SVN	-0.163	-0.186	-0.002	-0.002	-0.003	-0.003	-0.004	-0.006	-0.004	-0.006

Taula 3.4.6. Comparació entre els experiments 2 i 4.

Com s'observa a la taula 3.4.6, en quasi tots els mètodes es pot afirmar que els resultats no tenen cap variació significativa, a excepció de l'algoritme kNN. Tot i no ser l'algoritme que millors resultats dona, fa la funció d'indicador de la millora del model. Així doncs, sembla ser que l'aproximació utilitzada en l'experiment 4 resulta funcionar millor que la realitzada en el segon experiment.

Un cop vista la comparació, els algoritmes que millor funcionen segueixen sent els mateixos que en l'experiment 2: DT i AB. La regressió lineal torna a donar els mateixos problemes, tot i que s'obtenen valors molt alts durant les diferents repeticions, arriba un punt que en les monedes amb menys dades no troba cap model i dona valors baixos.

En la columna de paràmetres es poden veure els paràmetres utilitzats pels algoritmes més repetits durant els experiments. En cada experiment s'han optimitzat els diferents paràmetres de cada algoritme per tal d'obtenir el valor d' R^2 més alt possible. El fet de que els paràmetres no siguin constants en cada experiment indica que, si fossin fixats, els resultats serien més baixos.

Comparant els resultats obtinguts amb l'aproximació dels valors borsaris amb els del tercer experiment, es pot concloure que els dos conjunts de dades (monetari i borsari), tenen la mateixa importància, donat que els resultats no varien pràcticament. En el cas de les monedes amb més dades l'algoritme kNN dona valors més alts, mentre que en les monedes restants (Ripple i lota) els resultats són més baixos. Per la resta d'algoritmes els resultats pràcticament no varien.

4.5. Experiment 5

En aquest experiment s'utilitzaran com a atributs dades relatives a les criptomonedes, a borses de diferents països i a monedes de diferents països. El conjunt de dades està format pels mateixos valors relatius a les criptomonedes que en els experiments anteriors, afegint-hi els valors dels índexs borsaris de l'experiment 4 i els de les monedes de l'experiment 3.

D'aquestes dades es seleccionen com a atributs les mateixes que en el quart experiment afegint-hi els valors de les 32 monedes de l'experiment 3

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Regressió lineal (LR)
- K veïns més propers (kNN)
- Arbres de decisió (DT)
- *Adaptive boosting* (AB)
- Màquines de vectors de suport (SVM)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les dades de manera aleatòria els valors d' R^2 obtinguts variaran. El conjunt d'entrenament estarà format per aproximadament el 60% de les dades, i es validarà amb les dades restants.

En les taules 3.5.1 – 3.5.5, es poden observar els resultats obtinguts separats per criptomoneda:

Bitcoin	R^2		Temps (s)	Paràmetres
LR	0.997	± 0.0008	0.0262	N/A
kNN	0.900	± 0.0138	1.4206	K={11, 5}, pesos={distància}
DT	0.994	± 0.0114	0.0530	N/A
AB	0.991	± 0.0019	7.8192	10, 75
SVM	-0.148	± 0.0240	15.7308	kernel=rbf

Taula 3.5.1. 1172 dades amb un grup d'entrenament de 700 dades.

Litecoin	R^2		Temps (s)	Paràmetres
LR	0.992	± 0.0134	0.0268	N/A
kNN	0.850	± 0.0164	1.6660	K={3, 7}, pesos={distància}
DT	0.992	± 0.0020	0.1560	N/A
AB	0.993	± 0.0013	6.5268	25, 50
SVM	0.000	± 0.0023	15.8766	kernel=rbf

Taula 3.5.2. 1172 dades amb un grup d'entrenament de 700 dades.

Ethereum	R^2		Temps (s)	Paràmetres
LR	0.991	± 0.0133	0.0326	N/A
kNN	0.882	± 0.0148	0.7532	K={7, 5}, pesos={distància}
DT	0.992	± 0.0005	0.0362	N/A
AB	0.991	± 0.0016	5.3586	50, 100
SVM	0.000	± 0.0037	6.6782	kernel=rbf

Taula 3.5.3. 778 dades amb un grup d'entrenament de 450 dades.

Ripple	R^2		Temps (s)	Paràmetres
LR	0.939	± 0.0784	0.0218	N/A
kNN	0.811	± 0.0576	0.3358	K={3, 5}, pesos={uniforme, distància}
DT	0.964	± 0.0179	0.0256	N/A
AB	0.976	± 0.0185	3.4922	50, 100
SVM	-0.007	± 0.0048	1.5220	kernel=rbf

Taula 3.5.4. 342 dades amb un grup d'entrenament de 200 dades.

lota	R^2		Temps (s)	Paràmetres
LR	0.958	± 0.0678	0.0200	N/A
kNN	0.821	± 0.0421	0.3350	K={3}, pesos={distància}
DT	0.971	± 0.0087	0.0214	N/A
AB	0.980	± 0.0053	3.3764	50, 75
SVM	-0.009	± 0.0091	1.4588	kernel=rbf

Taula 3.5.5. 317 dades amb un grup d'entrenament de 200 dades.

A continuació es procedeix a comparar els resultats obtinguts en aquest experiment amb el que millors resultats a donat fins ara, l'experiment 4:

	Bitcoin		Litecoin		Ethereum		Ripple		lota	
	Exp.4	Exp.5	Exp.4	Exp.5	Exp.4	Exp.5	Exp.4	Exp.5	Exp.4	Exp.5
LR	0.997	0.997	0.997	0.992	0.995	0.991	0.922	0.939	0.771	0.958
kNN	0.892	0.900	0.865	0.850	0.862	0.882	0.839	0.811	0.809	0.821
DT	0.994	0.994	0.990	0.992	0.992	0.992	0.964	0.964	0.977	0.971
AB	0.992	0.991	0.994	0.993	0.990	0.991	0.982	0.976	0.982	0.980
SVN	-0.180	-0.146	-0.002	-0.000	-0.003	-0.000	-0.006	-0.007	-0.006	-0.009

Taula 3.5.6. Comparació entre els experiments 4 i 5.

Com es pot veure a la taula 3.5.6, els resultats són pràcticament els mateixos tret de la important millora de la regressió lineal en el cas de la lota. D'altra banda, fent servir un altre cop l'algoritme kNN com a indicador, es pot arribar a la conclusió de que aquest nou model millora una mica respecte als utilitzats en la resta d'experiments.

En la columna de paràmetres es poden veure els paràmetres utilitzats pels algoritmes més repetits durant els experiments. En cada experiment s'han optimitzat els diferents paràmetres de cada algoritme per tal d'obtenir el valor d' R^2 més alt possible. El fet de que els paràmetres no siguin constants en cada experiment indica que, si fossin fixats, els resultats serien més baixos.

4.6. Resum dels experiments 1-5

Un cop realitzats els cinc experiments amb la finalitat de determinar quin era el millor model i quin algoritme d'aprenentatge autònom donava millors resultats, s'ha arribat a la conclusió que els algoritmes que millors resultats donen són el DT i l'AB.

A continuació es mostren una sèrie de taules amb els resultats de cada algoritme utilitzats per a cada experiment:

Regressió lineal					
R^2					
	Bitcoin	Litecoin	Ethereum	Ripple	Iota
Experiment 1	0.998	0.998	0.998	0.993	0.993
Experiment 2	0.997	0.995	0.974	0.962	0.472
Experiment 3	0.998	0.998	0.997	0.992	0.991
Experiment 4	0.997	0.997	0.995	0.922	0.771
Experiment 5	0.997	0.992	0.991	0.939	0.958

Taula 3.6.1. Resultats obtinguts utilitzant la Regressió lineal.

Com es pot veure a la taula 3.6.1, tot i tenir uns resultats molt alts en la majoria dels casos, la regressió lineal falla en certs experiments donant resultats baixos. És per això, que s'ha descartat l'ús de la regressió lineal de cara als següents experiments.

K veïns més propers					
R^2					
	Bitcoin	Litecoin	Ethereum	Ripple	Iota
Experiment 1	0.963	0.801	0.777	0.925	0.952
Experiment 2	0.874	0.774	0.791	0.807	0.717
Experiment 3	0.858	0.779	0.769	0.928	0.948
Experiment 4	0.892	0.865	0.862	0.839	0.809
Experiment 5	0.900	0.850	0.882	0.811	0.821

Taula 3.6.2. Resultats obtinguts utilitzant kNN.

En la taula 3.6.2 es pot observar com els valors obtinguts amb l'algoritme kNN no són tant alts com els obtinguts amb altres algorismes. A aquest fet s'afegeix la variació dels resultats entre criptomonedes i experiments. Tot i així es pot fer servir com a indicador per tal de determinar quin dels models proposats funciona millor.

Arbres de decisió					
R^2					
	Bitcoin	Litecoin	Ethereum	Ripple	Iota
Experiment 1	0.995	0.993	0.993	0.979	0.981
Experiment 2	0.995	0.993	0.991	0.974	0.983
Experiment 3	0.994	0.993	0.991	0.957	0.980
Experiment 4	0.994	0.990	0.992	0.964	0.977
Experiment 5	0.994	0.992	0.992	0.964	0.971

Taula 3.6.3. Resultats obtinguts utilitzant arbres de decisió.

Adaptative boosting					
R^2					
	Bitcoin	Litecoin	Ethereum	Ripple	Iota
Experiment 1	0.992	0.992	0.991	0.978	0.980
Experiment 2	0.992	0.994	0.990	0.975	0.983
Experiment 3	0.992	0.992	0.991	0.971	0.980
Experiment 4	0.992	0.994	0.990	0.982	0.983
Experiment 5	0.991	0.993	0.991	0.976	0.980

Taula 3.6.4. Resultats obtinguts utilitzant AB.

Tant l'algoritme d'arbres de decisió com l'*adaptative boosting* tenen pràcticament els mateixos resultats ja que l'AB està basat en arbres de decisió, per tant era d'esperar que els resultats dels dos algoritmes fossin molt semblants. Els resultats obtinguts amb l'ús d'aquests algoritmes són molt elevats, i a diferència de la regressió lineal no existeix cap variació significativa. Cal remarcar que amb les dues monedes de les que es disposa menys dades (Ripple i Iota) els resultats obtinguts són una mica més baixos que amb la resta de criptomonedes. Considerant aquests factors, i els resultats obtinguts amb la resta d'algoritmes, el DT i l'AB seran els seleccionats per a dur a terme els següents experiments.

Màquines de vectors de suport					
	R^2				
	Bitcoin	Litecoin	Ethereum	Ripple	Iota
Experiment 1	-0.180	0.014	0.001	-0.017	-0.006
Experiment 2	-0.163	-0.002	-0.003	-0.004	-0.004
Experiment 3	-0.164	0.004	0.001	-0.005	-0.031
Experiment 4	-0.186	-0.002	-0.003	-0.006	-0.004
Experiment 5	-0.148	-0.000	-0.000	-0.007	-0.009

Taula 3.6.5. Resultats obtinguts utilitzant SVM.

Com s'ha mencionat anteriorment, l'algoritme basat en màquines de vectors de suport, no ha funcionat per a cap dels casos, essent tots els resultats obtinguts molt propers a 0 i en la majoria dels casos amb signe negatiu.

4.7. Experiment 6

Un cop determinat el model i els algoritmes amb millors resultats, el proper pas es veure com es comporten al intentar predir els valors de les criptomonedes. Per tal de fer-ho s'utilitzaran les dades usades en el cinquè experiment. La principal diferència és que s'afegirà el valor actual de la criptomoneda a la llista d'atributs i el valor a calcular serà el preu futur de la criptomoneda. Es calcularà el valor de la criptomoneda a 1, 2, 3, 4, 5, 6 i 7 dies vista per tal de veure com evolucionen els valors d' R^2 en funció de la llunyania temporal.

Per tal de calcular el valor de les criptomonedes s'utilitzaran diferents algoritmes de regressió, majoritàriament de *Machine Learning*, continguts a la llibreria Scikit-Learn. Els utilitzats en aquest experiment són els següents:

- Arbres de decisió (DT)
- *Adaptive boosting* (AB)

L'efectivitat dels diferents algoritmes s'avaluarà emprant el valor de R^2 com a mesura d'avaluació i s'utilitzarà un protocol de validació simple. El protocol de validació simple consisteix en separar les dades en dos conjunts: *training* i *test*. Prèviament a separar les dades en els anteriorment mencionats grups, les dades es reordenaran de manera aleatòria, és per això que cada algoritme serà llançat un total de 5 vegades. D'aquesta manera, en les taules de resultats apareixerà la mitja de les 5 vegades que s'ha dut a terme l'experiment, ja que al reordenar les dades de manera aleatòria els valors d' R^2 obtinguts variaran. El conjunt d'entrenament estarà format per aproximadament el 60% de les dades, i es validarà amb les dades restants.

Bitcoin	R^2				Temps (s)	
	DT		AB		DT	AB
1 dia	0.986	± 0.0040	0.987	± 0.0022	0.0594	6.7374
2 dies	0.975	± 0.0055	0.981	± 0.0019	0.0640	7.2096
3 dies	0.975	± 0.0093	0.979	± 0.0049	0.0510	6.9244
4 dies	0.971	± 0.0149	0.982	± 0.0020	0.0508	6.8262
5 dies	0.958	± 0.0241	0.980	± 0.0023	0.0610	6.5138
6 dies	0.976	± 0.0033	0.982	± 0.0015	0.0535	6.9512
7 dies	0.961	± 0.0253	0.984	± 0.0040	0.0511	6.9528

Taula 3.7.1. Resultats obtinguts en la predicció del valor del Bitcoin.

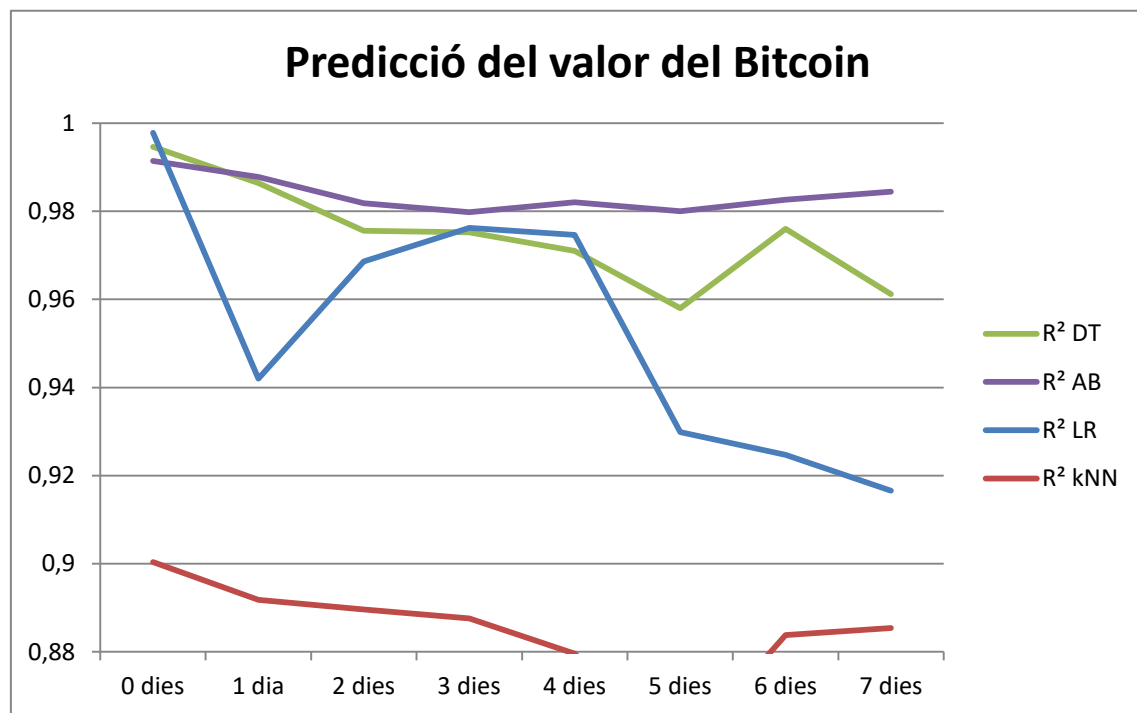


Figura 3.7.1. Resultats obtinguts en la predicció del valor del Bitcoin.

Tant a la taula 3.7.1 com a la figura 3.7.1 es poden observar els resultats d' R^2 obtinguts a dur a terme una predicció del valor del Bitcoin.

L'escala de la figura 3.7.1 s'ha adaptat per tal de veure la tendència dels dos millors algoritmes, és per això que alguns dels valors presos per l'algoritme kNN no es veuen a la gràfica. Es pot veure com els dos algoritmes seleccionats com a millors són els que millors funcionen a l'hora de predir el valor al llarg dels dies. Tot i que la regressió lineal sobrepassa en algun punt a l'algoritme basat en arbres de decisió, es pot veure com el comportament d'aquesta és molt més irregular.

La tendència del valor d' R^2 és la de disminuir al allunyar l'horitzó temporal de la predicció, tot i que ho fa de manera molt suau.

Litecoin	R^2				Temps (s)	
	DT		AB		DT	AB
1 dia	0.979	± 0.0087	0.984	± 0.0026	0.0744	6.7916
2 dies	0.970	± 0.0130	0.980	± 0.0066	0.0620	6.2540
3 dies	0.967	± 0.0161	0.975	± 0.0084	0.0962	5.7474
4 dies	0.966	± 0.0135	0.974	± 0.0046	0.0834	7.2282
5 dies	0.966	± 0.0223	0.964	± 0.0275	0.0786	6.7078
6 dies	0.970	± 0.0192	0.971	± 0.0104	0.0744	6.6104
7 dies	0.970	± 0.0146	0.966	± 0.0109	0.0769	6.5844

Taula 3.7.2. Resultats obtinguts en la predicció del valor del Litecoin.

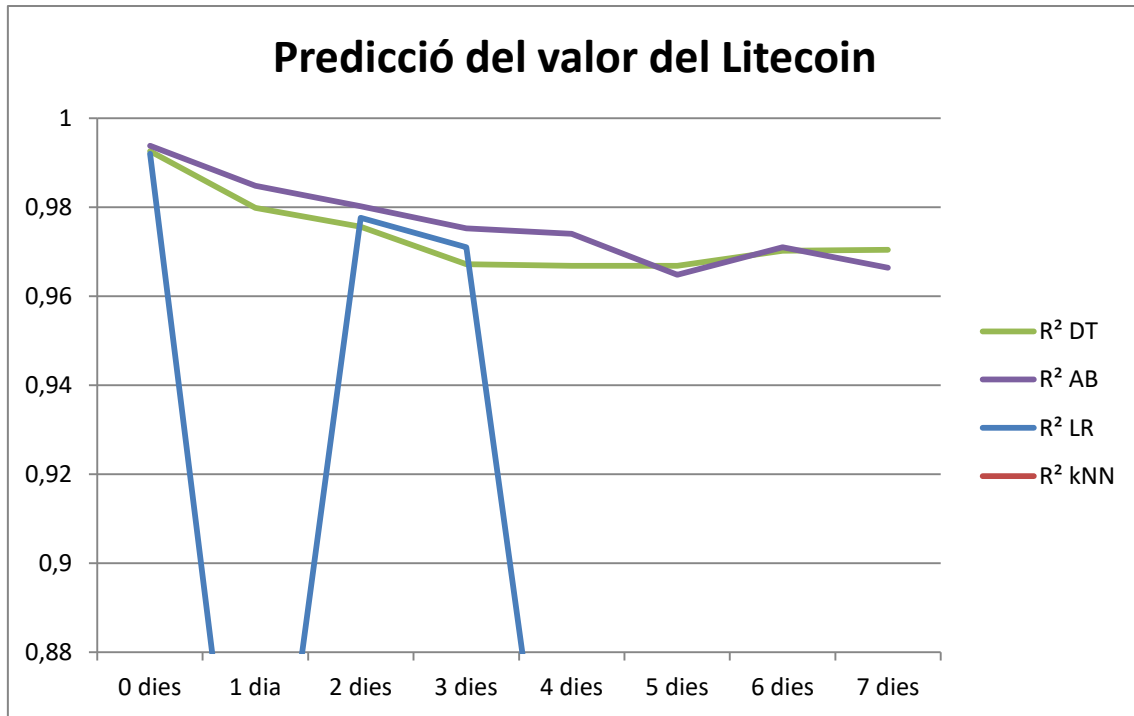


Figura 3.7.2. Resultats obtinguts en la predicció del valor del Litecoin.

Tant a la taula 3.7.2 com a la figura 3.7.2 es poden observar els resultats d' R^2 obtinguts a dur a terme una predicció del valor del Litecoin.

L'escala de la figura 3.7.2 s'ha adaptat per tal de veure la tendència dels dos millors algoritmes, és per això que alguns dels valors presos per la regressió lineal i els presos per l'algoritme kNN no apareixen a la gràfica. Cal destacar que la regressió lineal pren valors negatius, arribant fins a un mínim d' $R^2 = -0.666$.

Es pot veure com els dos algoritmes seleccionats com a millors són els que millors funcionen a l'hora de predir el valor al llarg dels dies i que la seva tendència és clarament descendent, tot i fer-ho de manera suau.

Ethereum	R^2				Temps (s)	
	DT		AB		DT	AB
1 dia	0.985	± 0.0013	0.986	± 0.0041	0.0446	6.0700
2 dies	0.975	± 0.0077	0.976	± 0.0058	0.0386	5.2852
3 dies	0.969	± 0.0128	0.978	± 0.0024	0.0384	5.1092
4 dies	0.970	± 0.0058	0.979	± 0.0029	0.0434	5.7116
5 dies	0.971	± 0.0062	0.976	± 0.0035	0.0450	5.3276
6 dies	0.966	± 0.0073	0.973	± 0.0064	0.0395	5.4564
7 dies	0.974	± 0.0054	0.975	± 0.0053	0.0522	5.2070

Taula 3.7.3. Resultats obtinguts en la predicció del valor de l'Ethereum.

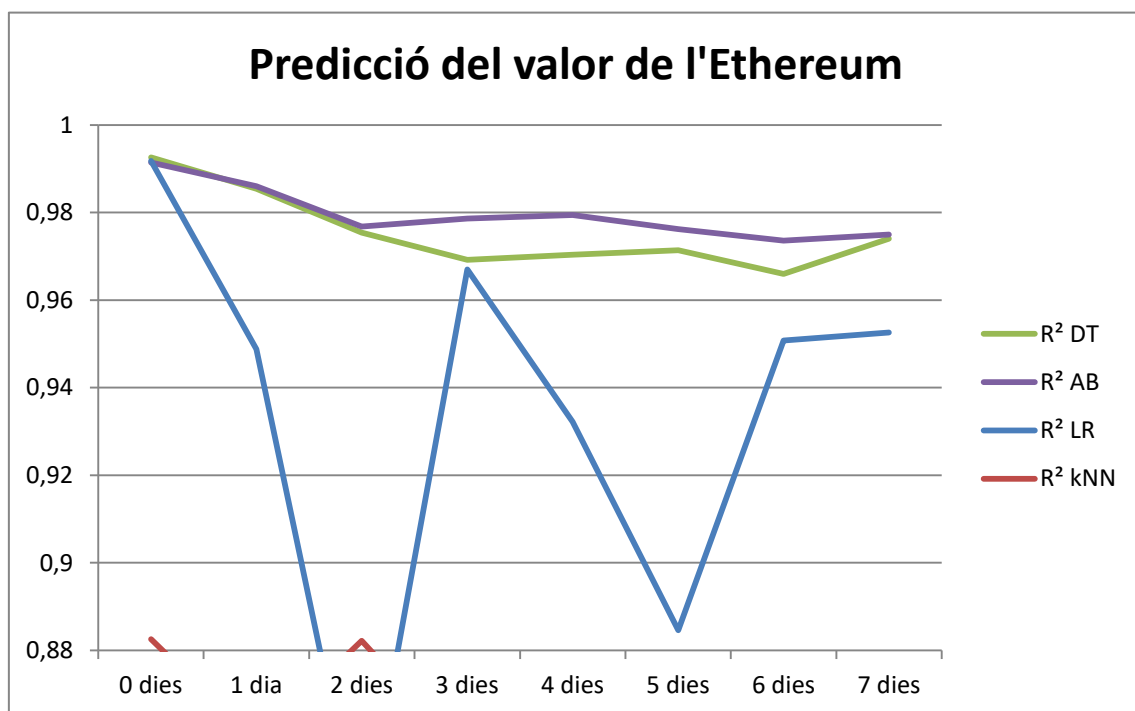


Figura 3.7.3. Resultats obtinguts en la predicció del valor de l'Ethereum.

Tant a la taula 3.7.3 com a la figura 3.7.3 es poden observar els resultats d' R^2 obtinguts a dur a terme una predicció del valor de l'Ethereum.

L'escala de la figura 3.7.3 s'ha adaptat per tal de veure la tendència dels dos millors algoritmes, és per això que alguns dels valors presos per la regressió lineal i els presos per l'algoritme kNN no apareixen a la gràfica. Tot i no aparèixer a la gràfica, els valors presos per l'algoritme kNN tendeixen a descendir d'una manera molt semblant als algoritmes DR i AB, a diferència de la regressió lineal, que com en els altres casos, té un comportament molt irregular.

Es pot veure com els dos algoritmes seleccionats com a millors són els que millors funcionen a l'hora de predir el valor al llarg dels dies i que la seva tendència és clarament descendent, tot i fer-ho de manera suau.

Ripple	R^2				Temps (s)	
	DT		AB		DT	AB
1 dia	0.953	± 0.0177	0.964	± 0.0025	0.0252	3.4188
2 dies	0.906	± 0.0547	0.933	± 0.0241	0.0230	3.1652
3 dies	0.896	± 0.0565	0.907	± 0.0531	0.0286	3.1400
4 dies	0.880	± 0.0799	0.922	± 0.0444	0.0238	3.0480
5 dies	0.893	± 0.0939	0.942	± 0.0116	0.0240	2.9968
6 dies	0.921	± 0.0210	0.9582	± 0.0158	0.0279	3.0958
7 dies	0.926	± 0.0251	0.940	± 0.0143	0.0251	3.3234

Taula 3.7.4. Resultats obtinguts en la predicció del valor del Ripple.

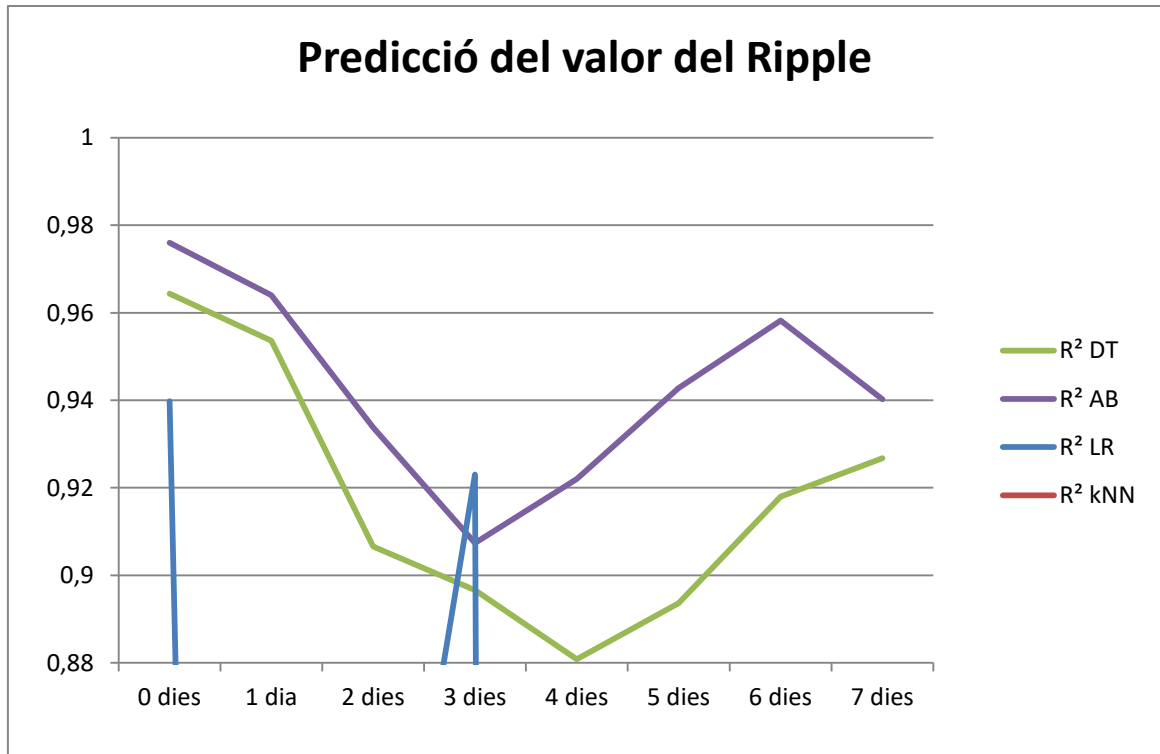


Figura 3.7.4. Resultats obtinguts en la predicció del valor del Ripple.

Tant a la taula 3.7.4 com a la figura 3.7.4 es poden observar els resultats d' R^2 obtinguts a dur a terme una predicció del valor del Ripple.

L'escala de la figura 3.7.4 s'ha adaptat per tal de veure la tendència dels dos millors algoritmes, és per això que alguns dels valors presos per la regressió lineal i els presos per l'algoritme kNN no apareixen a la gràfica. Cal destacar que la regressió lineal pren valors negatius, arribant fins a un mínim d' $R^2 = -4.107$.

A diferència dels tres casos anteriors, la tendència dels dos millors algoritmes no tendeix a descendir amb la mateixa suavitat. Es pot observar com experimenta una caiguda brusca per al 3r i 4t dia i posteriorment torna a recuperar-se. Cal destacar també, que en cap dels tres casos anteriors s'havia baixat per sota d' $R^2 = 0.964$, i en aquest cas s'arriba a valors d' $R^2 = 0.88$.

lota	R^2				Temps (s)	
	DT		AB		DT	AB
1 dia	0.946	± 0.0142	0.958	± 0.0127	0.0258	3.4744
2 dies	0.930	± 0.0159	0.947	± 0.0108	0.0248	3.0364
3 dies	0.930	± 0.0252	0.948	± 0.0112	0.0238	3.2710
4 dies	0.936	± 0.0295	0.952	± 0.0277	0.0262	3.5232
5 dies	0.915	± 0.0494	0.922	± 0.0421	0.0630	3.2560
6 dies	0.912	± 0.0782	0.8876	± 0.0837	0.0240	3.2156
7 dies	0.928	± 0.0386	0.9282	± 0.0549	0.0246	3.5656

Taula 3.7.5. Resultats obtinguts en la predicció del valor de la lota.

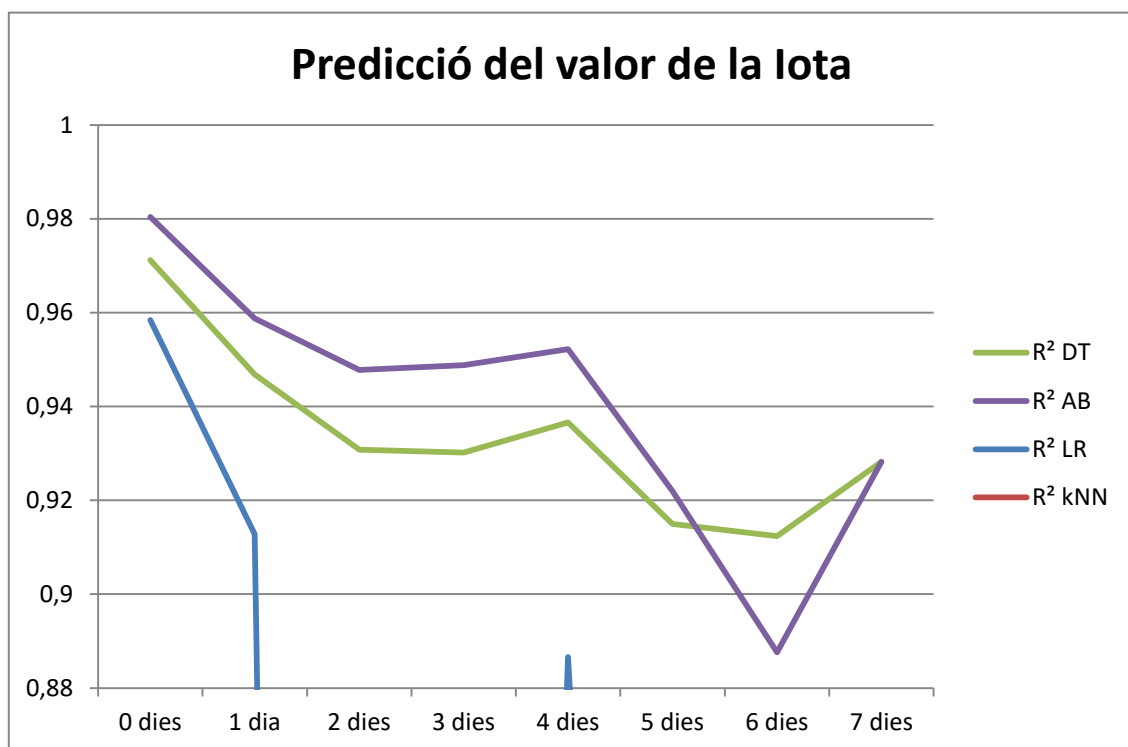


Figura 3.7.5. Resultats obtinguts en la predicció del valor de la lota.

Tant a la taula 3.7.5 com a la figura 3.7.5 es poden observar els resultats d' R^2 obtinguts a dur a terme una predicció del valor de la lota.

L'escala de la figura 3.7.5 s'ha adaptat per tal de veure la tendència dels dos millors algoritmes, és per això que alguns dels valors presos per la regressió lineal i els presos per l'algoritme kNN no apareixen a la gràfica. Cal destacar que la regressió lineal pren valors negatius, arribant fins a un mínim d' $R^2 = -2.518$.

De la mateixa manera que en el cas del Ripple, la tendència de descens de l' R^2 en el cas de la Lota, és bastant més brusca que en les tres primeres criptomonedes. Tant aquest fet, com el fet d'obtenir resultats més baixos, s'associa a la diferència en volum de dades entre les diferents monedes.

5. Anàlisi de l'impacte ambiental

En el present projecte, al tractar-se d'un estudi de viabilitat, no es generen canvis susceptibles directes per al medi ambient, així com cap mena de residu contaminant. Tot i això cal mencionar que el dispositiu emprat en aquest projecte (ordinador portàtil) consumeix electricitat, i està compost per elements que en el seu procés d'obtenció sí que influeixen negativament en el medi ambient.

Tot i l'indicat en el paràgraf anterior, es considerarà negligible l'impacte ambiental d'aquest projecte.

Conclusions

Mitjançant aquest projecte es pretenia determinar si era o no viable crear una aplicació o programa de compra/venta automàtica de criptomonedes. Per tal de donar resposta a aquesta qüestió s'han hagut d'assolir altres objectius menors prèviament.

En primer lloc, es pot afirmar que l'extracció de dades s'ha assolit completament. Aquest era un requisit indispensable, ja que sense dades no hagués sigut possible dur a terme els següents passos del projecte.

En segon lloc, aquestes dades han hagut de ser tractades abans de ser introduïdes als algorismes d'aprenentatge automàtic. Aquest ha sigut el punt amb més complexitat del projecte donat que s'han manejat una gran quantitat de dades. Per tal d'assolir aquest objectiu ha sigut necessari un procés d'aprenentatge previ, ja que amb els coneixements assolits durant el grau no era possible dur a terme cap dels dos primers objectius.

En tercer lloc, es pot afirmar que la previsió del preu de les criptomonedes ha estat un èxit. Si bé és cert que els resultats han estat millor amb les monedes Bitcoin, Litecoin i Ethereum, fet que s'associa amb el volum de dades disponible, els resultats de les monedes Ripple i Lota segueixen sent molt bons.

Els resultats obtinguts al llarg d'aquest treball han estat molt prometedors, tot i això, abans de començar amb el desenvolupament de l'aplicació serà necessari dur a terme nous anàlisis.

Treballs futurs

Donat el temps disponible per dur a terme aquest projecte, existeixen una sèrie de treballs que no s'han pogut dur a terme. Per tal de completar aquest estudi, es proposen els següents treballs:

- **Estudiar el funcionament del model en els punts particulars**, és a dir, estudiar el l'efectivitat del model dissenyat en els punts on els actius experimenten pujades o baixades molt exagerades en el seu preu.
- **Estudiar si les diferències entre els mètodes són estadísticament diferents.**
- **Actualització automàtica de les dades.** Tot i que les dades són obtingudes i tractades de manera automàtica, actualment no s'actualitzen automàticament, és necessària l'execució dels scripts per a tal d'actualitzar tota la informació.
- **Predicció en funció de les notícies.** De manera complementaria a les tècniques emprades per predir el preu de les criptomonedes, es proposa la creació d'una eina que permeti fer una predicció de la tendència en funció de notícies o popularitat dels actius, per exemple Twitter.

Un cop duts a terme aquests treballs s'estaria llest per seguir amb el desenvolupament de l'aplicació.

Planificació

A continuació es mostra el diagrama de Gantt on es pot veure la planificació de les diferents tasques del projecte. Cal destacar el projecte es comença més tard del previst (febrer) degut a certes circumstàncies amb la matriculació del mateix.

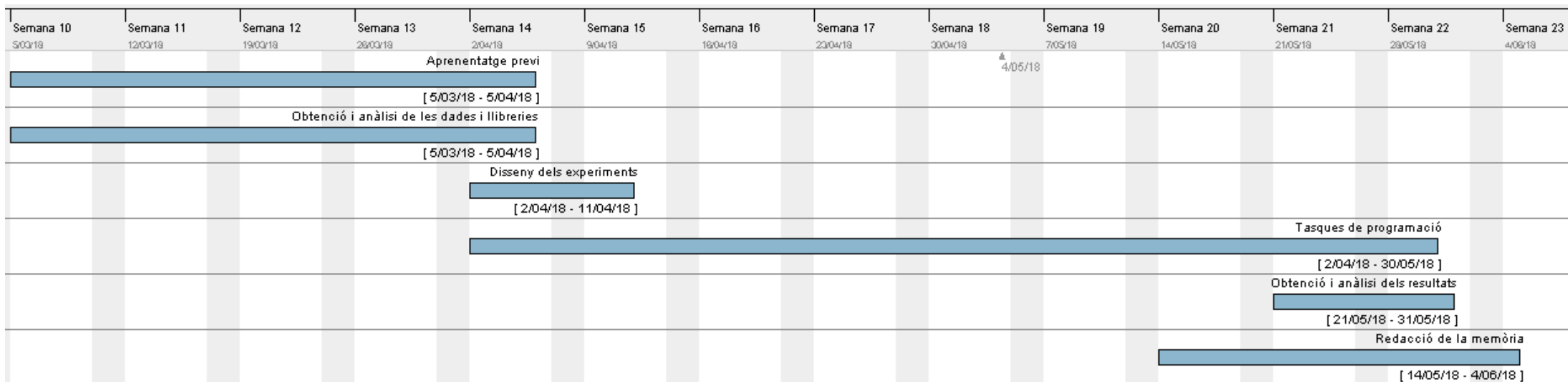


Diagrama de Gantt de la planificació del projecte.

Pressupost i/o anàlisi econòmica

En aquest apartat es mostren detalladament els costos associats al projecte.

Pressupost del prototip

El gran gruix dels costos ve donat per la mà d'obra. De les 605 hores invertides en la realització d'aquest estudi, la tasca que més temps ha ocupat ha sigut la programació dels scripts de tractament de dades, l'aprenentatge previ i predicció del valor de les criptomonedes.

Els costos materials són molt baixos, donat que només es fa servir un ordinador portàtil i no es té en compte el consum d'electricitat d'aquest, donat que és negligible respecte al cost total del projecte.

Les llibreries externes utilitzades, així com els recursos i eines utilitzades per a dur a terme aquest projecte no generen cap cost addicional, ja que es tracta de software lliure.

Concepte	Professional	Quantitat	Cost unitari	Cost
TASQUES				
Aprenentatge previ	Enginyer	135 h.	0 €/h	0 €
Recerca i anàlisi de dades/llibreries	Investigador	90 h.	20 €/h	1800 €
Disseny dels experiments	Enginyer	75 h.	35 €/h	2625 €
Tasques de programació	Programador	165 h.	30 €/h	4950 €
Obtenció i anàlisi dels resultats	Enginyer	60 h.	35 €/h	2100 €
Redacció de la memòria	Enginyer	80 h.	35 €/h	2800 €
MATERIAL				
Ordinador portàtil (Lenovo 300-15ISK)		1 u.	673 €/u	673 €
				TOTAL = 14948 €

Taula de costos de tots els elements del projecte.

Pressupost de producció

En aquesta secció no s'adjunta cap pressupost donat que l'aplicació tractada no està destinada a ser venuda ni comercialitzada donat que el mercat estudiat es tracta d'un sistema caòtic d'ordre 2. En el cas de que un gran número d'inversors es regís per les prediccions d'aquesta aplicació, aquestes deixarien de ser correctes, perquè en aquest tipus de sistemes el coneixement del futur modifica el mateix.

Bibliografia

- [1] STACKOVERFLOW. Developer community.
[<https://stackoverflow.com/questions/72738223>, 28 de maig de 2018]
- [2] PYTHON 3.6.5 DOCUMENTATION. Documentation for Python 3.6.5.
[<https://docs.python.org/3/>, 30 de maig de 2018]
- [3] EXPANSION. Diccionario económico.
[<http://www.expansion.com/diccionario-economico/macd.html>, 20 de maig de 2018]
- [4] BOLLINGER BANDS.
[<https://www.bollingerbands.com/>, 21 de maig de 2018]
- [5] STOCKCHARTS. Chart School.
[http://stockcharts.com/school/doku.php?id=chart_school:technical_indicators:relative_strength_index_rsi, 21 de maig de 2018]
- [6] CRYPTOCOMPARE. CryptoCompare API.
[<https://www.cryptocompare.com/api/#-api-data-pricehistorical->, 2 de juny de 2018]
- [7] PYTHON SOFTWARE FOUNDATION. Google finance client 1.3.0.
[<https://pypi.org/project/googlefinance.client/>, 2 de juny de 2018]
- [8] PYTHON SOFTWARE FOUNDATION. Forex-python 1.0.0.
[<https://pypi.org/project/forex-python/>, 2 de juny de 2018]
- [9] RAUL E. LOPEZ BRIEGA. Matemáticas, análisis de datos y python.
[<https://relopezbriega.github.io/blog/2015/10/10/machine-learning-con-python/>, 27 de maig de 2018]
- [10] SCIKIT LEARN. Machine learning in Python.
[http://scikit-learn.org/stable/modules/linear_model.html, 3 de juny de 2018]
- [11] SCIKIT LEARN. Machine learning in Python.
[<http://scikit-learn.org/stable/modules/neighbors.html#regression>, 3 de juny de 2018]
- [12] SCIKIT LEARN. Machine learning in Python.
[<http://scikit-learn.org/stable/modules/tree.html#tree>, 3 de juny de 2018]
- [13] SCIKIT LEARN. Machine learning in Python.
[<http://scikit-learn.org/stable/modules/ensemble.html#adaboost>, 3 de juny de 2018]

[14] SCIKIT LEARN. Machine learning in Python.

[<http://scikit-learn.org/stable/modules/svm.html#svm-regression> , 3 de juny de 2018]

[15] COINMARKETCAP.

[<https://coinmarketcap.com/> , 4 de juny de 2018]

[16] Raúl Benítez, Gerard Escudero, Samir Kanaan, David Masip Rodó, *Inteligencia artificial avanzada*. Barcelona: Editorial UOC, 2013. ISBN 978-84-9029-887-9.